Checksum Validation (CPP-002)

CPP-Identifier	CPP-002
CPP-Label	Checksum Validation
Author	Kris Dekeyser
Contributors	Johan Kylander
Evaluators	Felix Burger, Maria Benauer, Laura Molloy
Date of edition completed	29.08.2025
Change history	**Comments**
Version 1.0 - 29.08.2025	Milestone version

1. Description of the CPP

The TDA validates checksums against those stored in the *Information Package* at *Ingest* or *Access*.

Inputs and outputs

Input(s)	
Data	*File*
Metadata	*Fixity metadata*
Output(s)	
Metadata	*Provenance metadata* (the checksum validation event, including the datetime)
Alerts	*File* corruption alert for each *File* that fails the check
	Unsupported checksum algorithm alert for each algorithm that is not supported by the system

Definition and scope

The Checksum Validation process compares the checksums that come with the *Information Package*, with checksums based on the current state of the *File*'s data. Checksums of *Files* are validated at *Ingest* or *Access* to a TDA. The checksums may be either part of 1) a *SIP*; 2) an imported *AIP*; or 3) an *AIP* stored in the TDA.

A misalignment between the checksum algorithms used in the *Information Package* and the TDA's checksum policy's list of algorithms may occur. For example, incoming *Information Packages* (*SIPs* and *AIPs*) may include a checksum with a different algorithm created by the producer. In the case of *AIPs*, the checksum policy might have been updated during the lifecycle of the TDA, resulting in the removal and/or addition of algorithms. In such cases, the process should use whatever algorithm is associated with a given checksum to the best of its potential.

Process description

Trigger event(s)

Trigger event	CPP-identifier
SIP or *AIP* import	CPP-029 (Ingest)
DIP generation	CPP-025 (Enabling Access)
AIP export	CPP-006 (AIP Batch Export)

Step-by-step description

No	Supplier	Input	Steps	Output	Customer
1	*SIP*	Fixity information	Get the list of pre-calculated checksums with their respective algorithms	List of checksums and algorithms	
	CPP-001 (Checksum Generation and Recording)				
2		List of checksums and algorithms	Evaluate for each algorithm individually if it is supported by the system	Algorithm supported (step 3)	
		Digital archive system configuration		Algorithm not supported: a) examine further procedure (e.g. based on legal agreements;	

				submission policies; communication with producer, if possible) b) Process completed	
3		*File*	For each algorithm, recalculate the checksum of the *File* and match it with the given checksum	All *File* checksums match (step 5)	
		List of checksums and algorithms		Alert that any of the *File* checksums does not match: a) Examine further procedure b) Process completed	
5		All *File* checksums match	Document the event and its timestamp	Datetime for the checksum generation in the *Provenance metadata*	

Rationale(s)[1] and worst case(s)

Rational	Impact of inaction or failure of the process
Making sure the data content is still in the condition as intended	Data corruption during transfers into the system or during export could go unnoticed

2. Dependencies and relationships with other CPPs

Dependencies

CPP-ID	CPP-Title	Relationship description
CPP-001	Checksum Generation	CPP-002 relies on fixity information as produced and stored by CPP-001, but only during CPP-025 Enabling Access and CPP-006 AIP Batch Export. During CPP-029 Ingest, the fixity information supplied by the *SIP* will be used instead.

Other relations

Relation	CPP-ID	CPP-Title	Relationship description
Required by	CPP-006	AIP Batch Export	To ensure the integrity of the data during transport from the TDA storage the exported *Files*' checksums need to be verified.
Required by	CPP-025	Enabling Access	As the *DIP* is created, all *File* checksums need to be validated to ensure that the *DIP* is representative of the *AIP.*
Required by	CPP-029	Ingest	For any *SIP* submitted to the TDA all included *File* checksums need to be checked to validate the integrity of the *Files.*

[1] Term derived from PREMIS.

Affinity with	CPP-004	Data Corruption Management	All new *AIP* copies must have their checksum validated to verify that the process was successful. However, the checksum validation is more mechanical in its nature, only aiming at verification of the copy process. The CPP-002 checksum validation is more comprehensive (including negotiations with producers and validation of results).
	CPP-030	Refreshment	
Affinity with	CPP-011	Replication	When *Files* are replicated, successful replication is validated by comparing the replicated *Files'* checksums against the original *Files'* checksums
Not to be confused with	CPP-003	Integrity checking	Both CPPs get input from CPP-001, and both calculate a checksum from an *Information package* and compare it to a given checksum. The difference is that CPP-002 is done during the *Ingest* or *Access* phases (relating to transfer of content, changes in space), while CPP-003 is done periodically during the preservation of the contents in the archival storage (relating to changes over time). Thus, CPP-002 and CPP-003 are not only triggered by different processes, but also trigger different responses.

3. Links to frameworks

Certification

Certification framework	Term used in framework to refer to the CPP	Section
CTS Link	Checksum (cf Extended Guidance documentation)/	Information Technology and Security/Storage & Integrity (R14)
Nestor Seal Link	/	/
ISO 16363 Link	/	/

Other frameworks and reference documents

Reference Document	Term used in framework to refer to the process	Section
OAIS Link	Quality assurance	4.2.3.3 (Ingest)
PREMIS Link	/	/

4. Reference implementations

Publicly available documentation

Institution	Organisation type	Language	Hyperlink
TIB – Leibniz Information Centre for Science and Technology and University Library, Germany	National library	English	https://wiki.tib.eu/confluence/spaces/lza/pages/93608391/Preservation+of+data+integrity+as+part+of+the+process+routines
	Non-commercial digital preservation service		
	Research infrastructure		
	Research performing organisation		
CSC – IT Center for Science Ltd., Finland	Non-commercial digital preservation service	Finnish	https://urn.fi/urn:nbn:fi-fe2024051731943 (Annex 3, section 2.1.1)
Archivematica	Digital preservation system	English	https://www.archivematica.org/en/docs/archivematica-1.17/user-manual/transfer/transfer/#create-a-transfer-with-existing-checksums